COARTICULATION IN CONTRASTIVE RUSSIAN STOP SEQUENCES

Lisa Davidson

Department of Linguistics, New York University, New York, USA
lisa.davidson@nyu.edu

ABSTRACT

The articulation of Russian stop-stop #CC, C#C, and #CəC is examined using ultrasound imaging. The tongue shape trajectories suggest that contrary to previous assumptions, C#C and #CC coarticulation and timing are not interchangeable. In some cases, native Russian #CC articulation is more similar to #CəC than to C#C, suggesting that learning the timing and coarticulation of these sequences may be a challenge for L2 acquisition.

Keywords: speech production, ultrasound, coarticulation, consonant clusters

1. INTRODUCTION[1]

Consonant sequences have been the focus of several articulatory studies examining coarticulation and gestural timing [2, 5, 8, 9]. For example, Recasens and colleagues have shown that gestural overlap between two stops is greatest at the midpoint of the closure period, and overlap occurs whether consonants are highly constrained lingually (e.g. /k/) or relatively unconstrained (e.g. /p/) [8, 9]. This research also showed that coarticulation is a matter of degree; for example, in Catalan, the alveolar fricative /s/ is more resistant to coarticulation than the dental stop /t/ [9].

The studies cited above provide data about a limited number of languages, mainly American [2, 8] and British English [5], and Catalan [9]. Additionally, the results are primarily from electropalatography and are based on either word-medial VCCV sequences or sequences containing a word boundary (C#C). In this study, previous findings are extended to an ultrasound study of three kinds of Russian stop-stop sequences: #CC, C#C, and #CəC. While Zsiga [11] provides acoustic information indicating that the first stop in C#C sequences is always released in Russian, her study neither includes #CC sequences nor focuses on spatial aspects of coarticulation. Thus, we compare the articulatory production of tautosyllabic Russian stop-stop sequences to other consonant sequences.

A second goal of this study is to determine how similar C#C sequences are to #CC sequences in a language which allows both of them. This question has ramifications for several issues in the literature: whether syllable structure affects coordination of consonant sequences [1, 2], whether coarticulatory resistance interacts with syllable structure [8, 9], and whether the ability to produce C#C sequences could usefully be transferred to the production of #CC sequences, which loanword studies have suggested may be unlikely [3, 10].

Cross-linguistic studies of coarticulation and articulatory timing lay a foundation for understanding what both first and second language (L2) learners need to acquire when learning consonant clusters. To shed light on the articulatory input for L2 acquisition, this study focuses on the production of native Russian speakers' #CəC sequences along with #CC and C#C. This is because we are ultimately interested in investigating how English speakers acquire Russian, and English speakers often insert a schwa-like vowel when producing non-native #CC sequences [3]. Thus, the third goal of this study is to examine the articulatory trajectories (produced by native speakers) that English learners of Russian must acquire to establish a phonological contrast between C#C, #CC and #CəC.

Ultrasound is well-suited to providing a holistic picture of tongue shape changes over time. We examine coarticulation by visualizing the movement of the whole midsagittal tongue surface from the onset of closure of C1 to the offset of closure for C2 for three consonant combinations (/kt/, /gd/, /tk/) and three sequences (#CC, C#C, and #CəC).

2. MATERIALS AND PROCEDURE

2.1. Participants

Four speakers were recruited from Russian communities in New York City. All speakers were born in Moscow and did not start learning English until they were between 12-17 years of age. These participants range in age from 20-29 years, and report speaking both English and Russian.

2.2. Materials

The target items in this study were three stop-stop sequences produced in the three contexts shown in Table 1. Because Russian exhibits word-final devoicing, C1 of the voiced sequence /g#d/ was devoiced, as indicated in Table 1 with a subscript circle under C1. Furthermore, because real phrases were used, the segments preceding the #CC and #CəC sequences were matched, but generally could not be for C#C. Accounting for context differences is discussed in Section 2.3. The following vowel, however, was matched for all three sequence types.

Table 1: Phrases containing target stimuli (in bold). The #CC words for /kt/ and /gd/ contain a clitic boundary between the consonants.

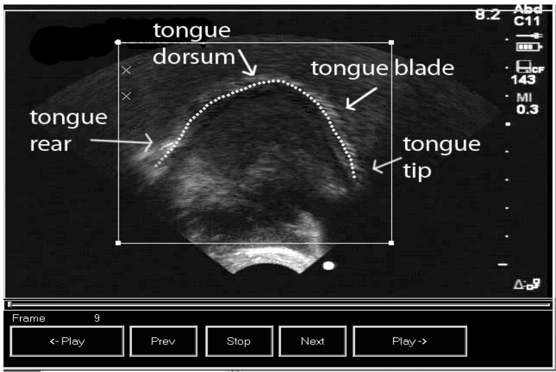
	C#C	#CC	#CəC
/kt/	[du'ra**k t**ɐk'sʲist] 'the taxi driver is an idiot'	[pəvʲer'nutʲ **kt**ɐk'sji] 'to turn to the taxi'	[sutʲ **kət**ɐ'strof] 'the essence of the catastrophes'
/gd/	[pɐ'bʲe**g̊ d**ɐ'moj] 'the run home'	[xotʲ **gd**ɐ'mam] 'at least to the houses'	[prətʃʲi'tatʲ **gəd**ɐ'voj] 'to read the annual'
/tk/	[nɐ'lʲi**t k**ɐ'nʲjak] 'the cognac is poured'	[cʲemʲ 'fabrʲik **tk**ɐnʲ'ja] '7 weaving factories'	[kak **tək**ɐ'va] 'as such'

2.3. Procedure

Midsagittal images of the tongue were recorded from a Sonosite Titan portable ultrasound machine using a 5-8MHz Sonosite C-11 transducer with a 90° field of view, a depth of 8.2cm, and a scan rate of 30 frames/sec. The ultrasound video signal and an audio signal were synchronized and captured directly into an avi file on a computer using a Canopus ADVC-1394 capture card and Adobe Premiere 6.0. Participants were seated in a sound-proof booth and their heads were stabilized using a moldable head stabilizer (Comfort Company) to ensure that images from different utterances could be compared. The transducer was held stable underneath the speaker's chin with a microphone stand. Each speaker produced each of the phrases in Table 1 10 times. The 10 repetitions of the phrases were presented in random order on a computer screen at eye level using Powerpoint.

Next, JPG image stills corresponding to the acoustic events of 10ms after onset of closure or frication of C1 to the release burst of C2 were extracted from the avi files. The 10ms mark was used to ensure that the tongue had already reached constriction and thus was less affected by coarticulation of the preceding phoneme. The JPG stills were loaded into EdgeTrak for measurement [6]. EdgeTrak is a computer program that automates the tracking of tongue contours by extracting (x,y)-coordinates from the lower edge of the white curve representing the tongue surface in the ultrasound image. 100 points were extracted for each tongue curve, which were then used for statistical analysis (see Figure 1).

Figure 1: Screenshot of the EdgeTrak extraction for a midsagittal tongue shape during the production of /g/.



2.4. Measurement

The consonant sequences are compared both through visual inspection and with a numerical similarity measure. For each frame, the tongue contours of the 10 individual repetitions are averaged and then displayed as a series of x, y, t surfaces using the program SURFACES [7]. These spatiotemporal figures are shown in Figures 2 & 3.

Next, the mean difference in millimetres along the length of the tongue curve for the averaged contours was calculated in SURFACES for the #CC~C#C, #CC~#CəC, and C#C~#CəC comparisons. Collapsing across all frames,[2] the absolute average differences across the entire consonantal sequence for each comparison were submitted to an ANOVA to determine differences in tongue shape over time in the comparison of #CC, C#C, and #CəC sequences.

3. RESULTS

3.1. Statistical analysis

A univariate ANOVA with the dependent variable of the difference in mean mm between tongue shapes averaged across all frames was performed. Subjects were treated as a random factor. The independent variables were consonant sequence type (/kt/, /gd/, /tk/) and comparison type

(#CC~C#C, #CC~#CəC, and C#C~#CəC). The mean absolute differences are shown in Table 2.

Results show a significant main effect of sequence type [F(2,6)=17.82, *p*<.003] and comparison type [F(2,6)=9.98, *p*<.02], and a significant interaction [F(4,12)=3.19, *p*=.05]. The main effect of subject was not significant [F(3,3)= 3.74, *p*=.16], nor were the interactions of subject and sequence [F(6,12)=1.45, *p*=.27] or comparison [F<1]. Pairwise comparisons indicating the significant differences are shown in Table 2.

Table 2: Absolute mean differences (in mm) by comparison and sequence type. Pairwise comparisons are in the bottom half of the table (‘ < ’ indicates a significantly smaller difference for that comparison)

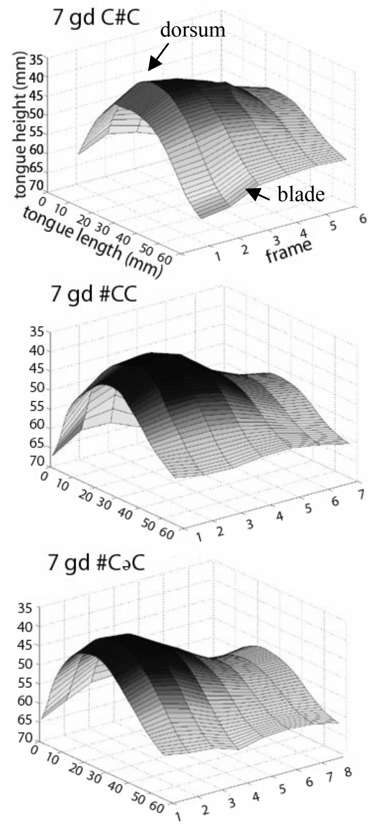
	#CC~C#C	#CC~#CəC	C#C~#CəC
/kt/	2.15	1.09	2.53
/gd/	1.38	1.03	1.62
/tk/	2.19	2.85	3.28
/kt/	#CC~#CəC < #CC~C#C, C#C~#CəC		
/gd/	#CC~#CəC < #CC~C#C, C#C~#CəC		
/tk/	#CC~C#C < #CC~#CəC, C#C~#CəC		

The interaction between sequence type and comparison shows that for /kt/ and /gd/, but not for /tk/, the difference between the onset cluster and the word boundary sequence (#CC~C#C) is significantly larger than the difference between the onset cluster and the schwa sequence (#CC~#CəC). This suggests that the timing and coarticulation of C#C sequences are not comparable to that of initial consonant sequences, and should not be used to make conclusions about tautosyllabic (or possibly word-medial) sequence. Possible causes for the greater similarity of #CC~#CəC and why #CC is more similar to C#C for /tk/ are discussed below.

3.2. Spatiotemporal visualization

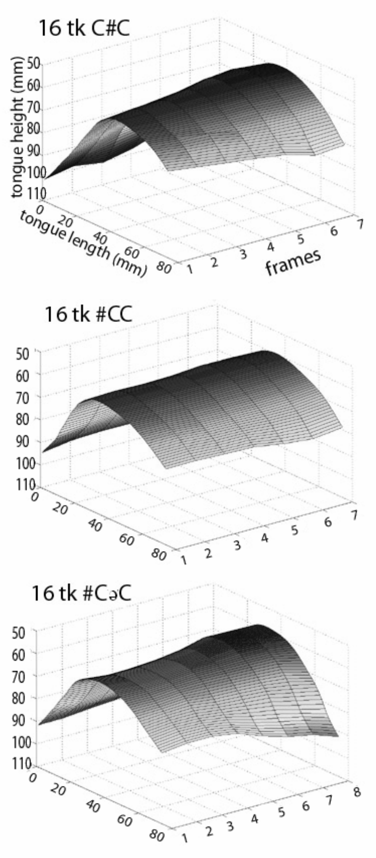
One difference between #CC and C#C is that the C1s are in different syllable/word positions. It has been shown that the articulation of onset consonants is more constricted and longer than coda articulations [1, 2].[3] Thus, one reason #CC is more similar to #CəC may be that C1 /k, g/ of the C#C sequence—a coda—has a less constricted dorsum position and different duration than onset velars do. As illustrated by speaker 7 in Figure 2, the constriction for the /g/ in both #CC and #CəC is similar in height, whereas the tongue dorsum position is lower in C#C (indicated by lighter gray). The lower position of the blade is also consistent with a generally lower tongue body position for C#C. The later frames for /d/ are similar for all sequence types, though coarticulation from the schwa results in a slightly lower tongue dorsum position for #CəC in the final two frames.

Figure 2: Articulations of all 3 sequences for speaker 7’s /gd/ series, connected for easier visualization of tongue shape changes. Darker blue/gray indicates a higher articulation (e.g., in the velar region), and yellow-orange/lighter gray indicates a lower tongue position (in the blade). (color on Image File 1)



The syllable position effects may also be interacting with the amount of coarticulatory resistance inherent to particular articulations [8, 9]. The ANOVA indicates an asymmetry between the sequences depending on whether C1 is velar or coronal. For /tk/, the onset cluster is more similar to C#C than to #CəC. One reason for this, illustrated by speaker 16 in Figure 3, is that because the dental /t/ may be less resistant to coarticulation (unlike the velar /k/) [9], its own tongue position is highly dependent on the following articulation. When followed by /ə/, it has a slightly lower dorsum position similar to the schwa’s, and when followed by /k/, the dorsum during /t/ is already raised. Thus, the trajectory of #CC is more similar to C#C.

Figure 3: Articulations of all three sequence types for speaker 16's /tk/ series (color on Image File 2)



4. DISCUSSION

4.1. All consonant clusters are not the same

These results suggest that syllable position effects may be mediated by whether or not a consonant is resistant to coarticulation. For non-resistant gestures, the influence of the surrounding articulations exerts a similar force regardless of context. For those that are more constrained, differences attributable to word or syllable position become evident. The distinction between different consonant types should be taken into account in studies of consonant cluster coarticulation.

4.2. Ramifications for language acquisition

The similarity of native Russian speakers' production of #CC and #CəC for some sequences implies that articulatory trajectories, in addition to any native phonological prohibitions, may contribute to difficulties in production and perception of consonant sequences by L2 learners. Because the articulation of #CəC can be more similar to #CC even for native speakers, it may be particularly difficult for learners to master the fine distinctions in articulation necessary to produce these sequences. Furthermore, even if a language contains C#C, a learner must discover that the articulation of C#C cannot be transferred to #CC.

5. REFERENCES

[1] Browman, C., Goldstein, L., 1995. "Gestural syllable position effects in American English," in *Producing Speech: Contemporary Issues for Katherine Safford Harris*, F. Bell-Berti and L. Raphael, Eds. New York: American Insitiue of Physics, pp. 19-33.

[2] Byrd, D., 1996. "Influences on articulatory timing in consonant sequences," *Journal of Phonetics* 24, 209-244.

[3] Davidson, L., 2006. "Phonology, phonetics, or frequency: Influences on the production of non-native sequences," *Journal of Phonetics* 34, 104-137.

[4] Fougeron, C., Keating, P., 1997. "Articulatory strengthening at the edges of prosodic domains," *Journal of the Acoustical Society of America* 101, 3728-3740.

[5] Hardcastle, W., Roach, P., 1979. "An instrumental investigation of coarticulation in stop consonant sequences," in *Current issues in the phonetic sciences*, H. Hollien and P. Hollien, Eds. Amsterdam: John Benjamins, pp. 531-540.

[6] Li, M., Kambhamettu, C., Stone, M., 2005. "Automatic contour tracking in ultrasound images," *Clinical Linguistics and Phonetics* 19, 545-554.

[7] Parthasarathy, V., Stone, M., Prince, J., 2006. "Spatiotemporal visualization of the tongue surface using ultrasound and kriging (SURFACES)," *Clinical Linguistics and Phonetics* 19, 529-544.

[8] Recasens, D., Fontdevila, J., Pallarès, M. D., Solanas, A., 1993. "An electropalatographic study of stop consonant clusters," *Speech Communication* 12, 335-355.

[9] Recasens, D., Pallarès, M. D., 2001. "Coarticulation, assimilation and blending in Catalan consonant clusters," *Journal of Phonetics* 29, 273-301.

[10] Ussishkin, A., Wedel, A., 2003. "Gestural motor programs and the nature of phonotactic restrictions: Evidence from loanword phonology," *WCCFL* 22, 505-518.

[11] Zsiga, E., 2003. "Articulatory timing in a second language: Evidence from Russian and English," *Studies in Second Language Acquisition* 25, 399-432.

---

[1] Thanks to Kevin Roon, Tuuli Adams, and Jason Shaw, and Serafina Shishkova for their input and assistance. This work was supported by NSF CAREER grant BCS-0449560.

[2] When the number of frames for each sequence was not the same, the number of frames corresponding to the shortest sequence was used. No time alignment was carried out, since differences in timing are crucial to this analysis and we did not want to obscure them.

[3] The syllable position differences may also be confounded by prosodic boundary effects [4].